

歴史的ドイツ語研究における コンピュータ利用の現状と可能性

新保 雅浩・高瀬 誠

はじめに

すでに10年前に筆者（新保）は古高ドイツ語研究のコンピュータ利用の一つとして古高ドイツ語作品Otfrids Evangelienbuchのインデックス（Wortindex zu Otfrids Evangelienbuch. Tübingen 1990）をドイツで出版した。この刊行に対して、ドイツはもとよりアメリカ、デンマーク、オランダ、フランス、スイス¹⁾などで書評という形で反響を得た。頂戴した様々な批判と要望には、新たな仕事によって応えなければならないと考え、ここ数年、画期的に飛躍したコンピュータ技術を駆使して、幾つかの試みを重ねてきた。そしてその仕事を完成するためには、歴史的ドイツ語（現代ドイツ語以外の古高ドイツ語、中高ドイツ語、初期新高ドイツ語を仮にこう呼んでおく）を研究するためにどんな情報が必要で、そのためにコンピュータは何が可能かをもう一度確認しておく必要があると感じた。たまたま筆者主催の中高ドイツ語輪読会に参加している高瀬氏がゲルマニ

1) スイスではチューリヒ大学S. Sonderegger教授が、フランスではソルボンヌ大学のY de Desportes教授が書評としてとり上げてくれた。また、ミュンスター大学のB. Meinke氏が詳細なご批判を相当数のページを割いて Zeitschrift für Dialektologie und Linguistik, LIX, Jahrgang, Heft1(1992)に寄せてくれた。尚、このインデックスの作成過程については以下の論文参照。

新保雅浩：古高ドイツ語におけるコンピュータ利用の一試み—Otfridのコンコーダント作成— 筑波大学言語文化論集第17号 1985

鈴木英一・新保雅浩：“Otfrid”のコンコーダントとその作成— コンピュータによる古高ドイツ語の研究に向けて— 文部省特定研究費「言語文化の理論的・実践的研究」所収 1990

ストであると同時に情報処理授業科目の講師でもあることから、今回共同でこの問題に取り組むことにした。1部「歴史的ドイツ語研究とコンピュータ利用の現状」、「歴史的ドイツ語研究にどんな情報が必要か」、という問題については主として新保雅浩が担当し、2部「歴史的ドイツ語研究に必要な情報はコンピュータ利用によってどこまで可能か、またそのためにどのような作業が必要か」というコンピュータの技術的問題については高瀬誠が担当し、後で互いに議論してより良い方法を導き出すべく努力した。

第I部

新保 雅浩

1. 歴史的ドイツ語研究とコンピュータ利用の現状

1.1 テキスト読解の為の道具

歴史的ドイツ語研究の第一歩は当然、古高ドイツ語や中高ドイツ語で書かれたテキストを読むこと、である。しかし現代ドイツ語のテキストを読むのとは大分事情が異なる。現代ドイツ語が規範化された共通語であり、その為、数多くの文法書や辞書の類があるのに対して、歴史的ドイツ語は共通語ではない。たとえば古高ドイツ語、中高ドイツ語という共通語が存在するわけではなく、古高ドイツ語、中高ドイツ語とはドイツ語史の中でそれぞれ一定の言語的特徴を共有する時代区分であるに過ぎない。従ってそれぞれのテキストがそれぞれの方言性を色濃く保持しているのが実状で、古高ドイツ語とは方言の一種の集合体と解することが出来る。さらに現代ドイツ語との大きな相違はこれらの歴史的ドイツ語にはいわゆるネーティヴスピーカーが存在しない点である。

それゆえ歴史的ドイツ語のテキストを読むためには現代語の文法書や辞書の欠如を補い、さらにネーティヴの代替となるべき道具立てが必要となる。一つのテキスト読解には、そのテキストに関するいくつかの写本、校訂本、グロサール、コメンタール、そのテキスト専用の辞書および文法書、さらには現代語翻訳など先達によって蓄積された研究成果を道具立てにしなければならない。すなわち、まず文献学的方法によるテキスト理解がこの分野の研究の原点となり、日本の現状としてはテキスト理解が研究の大部分をも占めてしまうのである。

古高ドイツ語の具体例を示そう。古高ドイツ語で書かれた作品の中でおそらくOtfrids Evangelienbuch（「オトフリートの福音書」）（以下Otfrid）が最も研究条件の整っている作品であろう。Otfridに関して蓄積された研究成果は古いドイツ語すなわち歴史的ドイツ語における言語研究のモデルともなるので、少し詳しく文献について述べておこう。

Otfridは860年頃にヴァイセンブルクの学僧オトフリートによって南ラインフランク語で書かれた共観福音書である。当時の文献の多くがラテン語からの翻訳であり、かつ断片のみ残されている中で、Otfridは作者オトフリートの創作であり、完全な写本の形で残っている。さらに作者自身が写本を校正した事実が判明したこともあって、何よりも写本状態が理想的である。古高ドイツ語時代で最も言語的な研究状況がよい文献といわれる条件をよく備えている（それゆえ内容面での研究も進んでいるわけである）。ここではテキストを読むための言語的道具立てとしての研究成果に限定して言及しよう。Otfridを出来るだけ正確に読むためにはP.Piper, J.Kelle, O. Erdmannという三人のOtfrid研究者の業績が特に重要である。

P.PiperはOtfridの校訂本、Otfrid専用のグロサールと文法書を残してい

2) Piper, Paul : Otfrids Evangelienbuch mit Einleitung, erklärenden Anmerkungen ausführlichem Glossar und einem Abiss der Grammatik. Freiburg 1887

る²⁾。このグロサルはインデックス、コンコーダンスをも兼ねており、原則として出現するすべての語彙がアルファベット順に配列され、さらに語義だけでなく、出現箇所すべての語形に関して文法形態がすべて記されている。すなわち動詞の場合は、話法、時称、数、人称形が、名詞、形容詞、代名詞の場合は、性、数、格形が表示されている。難点は追い込み印刷であるので検索に時間がかかること、欠落、誤植が見られることである。当時の手作業を考えれば当然のこととも言えるが、本書が現在絶版となっていることもあり、Otfrid理解を一層深めるためにも本書の改訂は必須といえる。

P.Piperのグロサルを補完するのがJ.Kelleの詳細なグロサルである³⁾。P.Piperのそれが文法形態の指示に詳しいのに対して、J.Kelleのそれは各語彙の用法に関する記述の詳しい辞書である。しかしこれもよく調べると欠落や誤植が散見される。さらにO.Erdmann編の校訂本⁴⁾がaltdeutsche Textbibliothekのシリーズに入っているが、巻末に語彙集があり、適切な語義を示している。さらにOtfridにはJ.KelleとP.Piper以外にO.ErdmannによるOtfrid専用のシンタックスとコメンタールがある⁵⁾。ちなみにO.ErdmannのDeutsche Syntax⁶⁾の古高ドイツ語の用例はほとんどOtfridからのものである。これだけの道具立てを駆使して我々はOtfridを少なくとも文字面だけは理解できるようになるのである⁷⁾。一般的な古高ドイツ語辞典（たとえばR.SchützeichelのAlthochdeutsches WörterbuchやE.G.GraffのAlthochdeutscher Sprachschatz）ではOtfridは読めない。さらに

3) Kelle, Johann : Otfrids von Weissenburg. Evangelienbuch. Text Einleitung Grammatik Metrik Glossar, Aalen 1963

4) Erdmann, Oskar : Otfrids Evangelienbuch. Halle 1882

5) Erdmann, Oskar : Untersuchungen über die Syntax der Sprache Otfrids. Olms 1973

Erdmann, Oskar : Otfrids Evangelienbuch. Herausgegeben 5. Auflage besorgt von Ludwig Wolff 1965

6) O.Erdmann : Grundzüge der deutschen Syntax nach ihrer geschichtlichen Entwicklung. Stuttgart 1886

7) その他にOtfridに関する次のような辞書類が出版されている。

Köbler, Gerhard : Verzeichnis der Übersetzungsgleichungen Otfrids von Weissenburg. Göttingen 1971

同時代の古高ドイツ語作品Tatianと古ザクセン語作品Heliandの道具立てを比較してみると、Otfridの研究成果の蓄積の厚みがよく分かる。TatianはOtfridと同様830年頃にラテン語から古高ドイツ語に翻訳された共観福音書である。Tatianを読むための道具立てとして主たるものはE.Sievers: *Tatian Lateinisch und althochdeutsch mit ausführlichem Glossar*(Nachdruck 1966)だけである。このグロサルは原則としてテキストの全ての語彙の文法形態（名詞の場合は性、格形、動詞の場合は話法、時称、数、人称形）の記述はかなり便利ではある。しかし初版以来、改訂がなされておらず、誤植、明らかな誤認、欠落が多い。それに何よりも目指す見出し語に達するにはかなりの時間がかかり、時には見つからず徒労に終わる。P.PiperやJ.Kelleに相当する辞書類はTatianにはないのである⁸⁾。Tatianは内容的にOtfridより分かりやすいことも道具立てが少ないことの一因であろう。

古ザクセン語で書かれたHeliandにも触れておこう。Heliandも上述の二つの古高ドイツ語作品と同じように共観福音書を内容とする9世紀中頃の作品である。古ザクセン語の代表的文献であるが、このテキスト専用のグロサルとしてはH.Seprtの *Vollständiges Wörterbuch zum Heliand und zur altsächsischen Genesis*. 1966 があり、テキストに出現するすべての語彙の文法形態記述はテキスト理解に便利である。Heliand専用のシンタックスもある（O.Behagehl: *Die Syntax des Heiland*. 1897）。古いドイツ語を読むためにはこのように必ずそのテキスト専用のグロサルやコメントールが必要なのである。しかし、ここでも道具立ての整っているテキストと整っていないテキストがある。またP.Piperのコンコーダント兼グロサルのように豊富な情報を含んでいながら、検索に時間がかかったり、検索し

8) 次のような辞書類も出版されている。

Köhler, Friedrich: *Lateinisch-althochdeutsches Glossar zur Tatianübersetzung*. Ferdinand Schöningh-Paderborn 1962

Köbler, Gerhard: *Verzeichnis der Übersetzungsgleichungen von Heliand und Genesis*. Göttingen 1971

にくいもの、あるいは誤植、欠落が改訂されないままのものがある。P.PiperのOtfridグロサールやE. SieversのTatianグロサールは改訂がなされれば最も便利な辞書の一つである。

古いドイツ語の読み方に関して言えば、中高ドイツ語も古高ドイツ語と同様に各テキストの為のインデックス、グロサール、コメンタールが必要となる。それゆえ 中高ドイツ語の多くの有名な作品にはすでに専用の道具立てが準備されている。Hartmann von Aueの代表作である宮廷叙事詩Iwein「イーヴァイン」に対しては G.F. Beneckeの辞書があり⁹⁾、中高ドイツ語時代最大の英雄叙事詩Nibelungenlied「ニーベルングンの歌」に対してもK.Bartschの三巻本中の一冊が専用の辞書¹⁰⁾ となっており、また中高ドイツ語後期に入ってもたとえばNithartの作品には E.Wieserによるナイトハルト専用の辞書¹¹⁾ がある。FrauenlobさえもK.StackmannのFrauenlob専用辞書¹²⁾ が作成されている。これらの研究成果の蓄積のある作品群には専用の辞書類がかなり早くから整っている。しかし特に中高ドイツ語後期あるいは初期新高ドイツ語の作品の中には研究成果の蓄積のない作品が数多く見られる。このような時代のテキストを読むときこそ、道具立ての必要性が生じるのである。たとえば、中高ドイツ語後期にWolkensteinの作品があるが、中高ドイツ語と新高ドイツ語的要素が混在しているドイツ語で、一般的な中高ドイツ語辞典だけでは読めない。 かつらうじてJ.Schatzの簡単なグロサール¹³⁾ があるが、絶版で印刷も悪く誤りも多い。かつてはこれだけの道具でWolkensteinの作品を読みこなすのは難しかったが、1973年にGöppingenでコンコーダントツが出版され¹⁴⁾、それ

9) Benecke, Georg: Wörterbuch zu Hartmanns Iwein. Dritte Ausgabe besorgt von C.B.Gorchling. Leipzig 1901
10) Bartsch, Karl :Der Nibelunge Not mit den Abweichungen von der Nibelunge Liet, den Lesarten sämtlicher Handschriften und einem Wörterbuche. Herausgegeben von Karl Bartsch. (1880) Georg Olms Hildesheim 1967
それ以外に、Lübben, August : Wörterbuch zu der Nibelunge Not(Liet). Wiesbaden 1960 がある。

11) Wieser, Edmund : Vollständiges Wörterbuch zu Neidharts Liedern. Leipzig 1989

12) Stackmann, Karl: Wörterbuch zur Göttinger Frauenlob-Ausgabe. Vandenhoeck & Ruprechtin Göttingen 1990

13) Schatz, Josef : Sprache und Wortschatz der Gedichte Oswalds von Wolkenstein. Wien 1930

は少なくとも単語の品詞判別に有効に働いた。それに同一テキストの他の用例を検索出来るだけでも有益であった。最近, Wolkensteinにコメンタールが著されたが¹⁵⁾、これを基にテキストを読み直せば、さらに詳しいコンコーダントの作成が可能であろう。

さらに中高ドイツ語後期の Reinmar von Zweterの詩作品、Thomasin von Zerclaereの *der welsche gast*, Berhold von Regensburgの説教集、Renner, Fridank など、あるいは中高ドイツ語後期から初期新高ドイツ語に属する Wittenweilerの作品はわずかな注釈以外（Wittenweilerには現代語訳があるが）一体、どんな道具立てであろう。中高ドイツ語の詳しい辞書である BMZ(Benecke, Zarncke, Müller : *Mittelhochdeutsches Wörterbuch*)やいわゆる大レクサー(*Mittelhochdeutsches Handwörterbuch*)をひたすら調べ、見当づけるはかないのである。このような研究史の浅い、従って研究の蓄積がないテキストに関しては、三つの意味でインデックスやコンコーダントの作成が意味を持つ。

1. インデックスやコンコーダントによって、他のいくつかの用例を確認しながら読み進めることができる。なぜなら歴史的ドイツ語のテキスト理解においては、用例の頻度数がその箇所の用法の正当性を確認する唯一の手段であるからである。

2. 読解の成果として、逐一インデックスにその語彙の意味や文法的形態を書き加えることによって、それ自体がテキスト理解の重要な情報として蓄積され、さらに詳しい道具立てへと発展する。

3. その作品専用の辞書編纂、一般的な中高ドイツ語の辞書編纂、あるいは既存の中高ドイツ語辞典の改訂などの為の基礎資料となる。

14) George Fenwick Jones / Hans-Dieter Mück / Ulrich Müller : *Verskonkordanz zu den Liedern Oswalds von Wolkenstein* Göttingen 1973

15) Marold, Werner: *Kommentar zu den Liedern Oswalds von Wolkenstein*. Bearbeitet und herausgegeben von Alan Robertshaw. Innsbruck 1995

1.2 最近のコンピュータ利用の実状

古高ドイツ語、中高ドイツ語、初期新高ドイツ語時代などの歴史的ドイツ語作品に関して、近年具体的にどのくらい新しい道具立てが加わったであろうか。すなわち Wolkenstein を例に述べたインデックスやコンコーダントの類が実際にどの位出版され、入手できる形になっているであろうか。その辺の事情について目立った動向を述べておきたい（辞書類に関しては今回は除外する）。

最近、コンピュータを利用して作成された歴史的ドイツ語作品のインデックスやコンコーダントが出版され始めている。ドイツの Max Niemeyer 出版社から出版されている Indices zur deutschen Literatur 「ドイツ文学作品のインデックスシリーズ」はその好例であろう。このシリーズ以外にも、アメリカでは中高ドイツ語文学作品に関して、インデックス類が早くから著されている。

* A. Senn and W. P. Lehmann : Word Index to Wolfram's Parzival. Madison 1938

* R.-M. S. Heffner and W. P. Lehman : A Word-Index to the Poems of Walther von der Vogelweide. Madison 1950

* M. E. Valk : Word-Index to Gotfried's Tristan. Madison 1958

また 1973 年に Ackermann aus Böhmen と Oswald von Wolkenstein の作品のインデックス

* Robert R. Anderson und James C. Thomas : Index Verborum zum Ackermann aus Böhmen. Amsterdam 1973

* George Fenwick Jones / Hans-Dieter Mück / Ulrich Müller : Verskonkordanz zu den Liedern Oswalds von Wolkenstein. Göttingen 1973
等が散発的に出版されている。その後、前述の Indices zur deutschen

Literaturのシリーズから古いドイツ語作品が現代ドイツ文学作品のそれに混じって出版されるようになった。現在まで既刊書を出版年月日順に挙げてみると

- * Hartmann von Aue: Lemmatisierte Konkordanz zum Gesamtwerk. 1979
- * Begriffsglossare und Indices zu Ulrich von Lichtenstein. 1980
- * Lemmatisierte Konkordanz zu den Schweizer Minnesängern. 1984
- * Wortindex zu Otfirids Evangelienbuch. 1990
- * Begriffsglossar und Index zu Ulrichs von Zatzikhoven Lanzelet 1993
- * Begriffsglossar und Index zur >Kudrun< 1994
- * Lemmatisierter Index zu den Werken des >Strickers< 1997
- * Wortindex zu hessisch-thüringischen Epen um 1200. 1997

現在までにおよそ30数冊出版されているシリーズの中で、古高ドイツ語、中高ドイツ語、初期新高ドイツ語時代のインデックスやコンコーダントはわずか8冊である。その内、古高ドイツ語が一冊 (Otfirids Evangelienbuch)、中高ドイツ語が6冊、初期新高ドイツ語はこのシリーズにはない。また特にここ10年の間にアメリカで、コンピュータを利用した大規模なコンコーダント類が出版されていることは注目に値する。

* Clifton Hall : A Complete Concordance to Wolfram von Eschenbach's Parzival. New York and London 1990

* Clifton Hall : A Complete Concordance to Gottfried von Strassburg's Tristan. Lewiston 1993

* Clifton Hall and Samuel Coleman : Walther von der Vogelweide. A Complete Reference Work. Head-Word and Rhyme-Word Concordances to His Poetry. Colorado 1995

* Clifton Hall : Head-Word and Rhyme- Word Concordances to des Minnesangs Frühling. Colorado 1997

コンピュータを利用したこれらの新しいインデックスあるいはコンコーダントはしかし編者の制作意図によって内容は必ずしも一様ではない。現在出版されているものを分類すると3つに大別することができる。

①純粋な語彙のインデックス（語彙のアルファベット順配列と出現箇所指示）

②コンコーダント（出現箇所表示とそのコンテキスト表示）

③語彙が概念別に分類されたコンコーダント

①のタイプは Wortindex zu Otfrids Evangelienbuch, Lemmatisierter Index zu den Werken des Strickers, Wortindex zu hessisch-thüringischen Epen um 1200のような、アルファベット順列（あるいは逆列）の語彙のインデックスを主としたものである。具体的には

aband [subst.] (1) 5-10-005

abande[subst.] (3) 3-14-055 4-02-007 4-11-011

abandes[subst.] (1) 5-04-009 (Otfrid)

これは全ての単語をそのままの語形でアルファベット順に並べ、品詞、頻度数、出現箇所の情報を与えている。あるいは

abend Subst Dbt 704

-abend: BR 143 105

-abende : VE xi 186

-abendes: KD 136 99

-abendes

-abent:KD (Stricker)

これは全ての単語の原形（名詞の場合は主格、動詞の場合は不定形）を見出し語にして、その見出し語をアルファベット順に並べて品詞を指示。その下位区分として各語形と出現箇所の情報を与えている。次の Wortindex zu hessisch-thüringischen Epen um 1200もこれに近い記述方法である。

abent (abend-) m AL(2) GL(1) HE(3) RU(1)

N Sg abunt HE(2)

- abut HE:12512

A Sg abint HE:7960s (218)

- abunt AL(2):6818 7168;HE(2):6721 7960

G Sg abendes RU:Ib 2

- abindis AT:C* 15

- abundes GL:977 ;HE (3):9584 12516 14990

②のタイプは①のタイプに比べると情報の質と量とが増大する。

Hartmann von Aue. Lemmatisierte Konkordanz zum GesamtwerkはHartmann von Aueの全作品〈Der arme Heinrich「哀れなハインリヒ」、Erec「エーレック」、Gregorius「グレゴリウス」、Iwein「イーヴァイン」、Das Klagebüchleinのコンコーダントである。これを例に説明しよう。

abent (11)

abendes (5)

381 E des abendes vil tiure

2474 E des abendes ze beiden siten

2485 E im was des abendes geschehen

4614 E des abendes do si gazen

2200 E des andern abendes gienc si dan

abent (5)

173 E unz daz der abent ane gie

1391 E den selben abent schouwen

273 I und do ez an den abent gienc

6489 I wand ez ein warmer abent was

691 K den abent und den morgen

abents (1)

789 I danne ouch des abents do ich reit

このコンコーダンスでは出現箇所 の頻度数と出現箇所 の引用があり、当該語彙の文脈をみることができる点が前者のそれと大きく異なる所である。Lemmatisierte Konkordanz zu den Schweizer Minnesängernはさらに当該語彙を大文字にして際だたせている。

ABENT

ABENT

UVSij.30,28 beidiu den ABENT unde fruo,

HVSix. 3,5 den ABENT den morgen

DVGxx.1,68 ABENT sehent, unde morgen,

JVWxxii,5,25 mangel tac, den ABENT und den morgen,

KVAXxiv.1,19 diu mich da twinget den ABENT den morgen ?

HAD xxvii,39,53 zuo zir, und ouch vil mangel ABENT spate.

ROSxxxii,7,16 ABENT unde morgen.

また、Minnesangs Frühlingは当該語彙をゴシックにして、中央に置いている。

abendes (1)

HM 136,30 set / do diu sunne, diu des abendes under

abent(4)

HM 135,4 /ich gelebte noch den lieben abent gerne./ daz si sich
her nider/

HM 146,(2) /swaz du tugende has./ den abent und den morgen/
sagent si

RH 175,13 waz. /gesache ich wider abent einen kleinen boten./so

RH 176,3 ist vil./spreche ich wider **abent** lihte ien schoene wort./
waz

③のタイプは語彙の語形を記述する以前に語彙を概念別に下位分類し、
その中で語彙をアルファベット順に配列したインデックスであり、
Begriffsglossare und Indices zu Ulrich von Lichtenstein、Begriffsglossar
und Index zur >Kudrun<、Begriffsglossar und Index zu Ulrichs von
Zatzikhoven Lanzeletがこのタイプである。Kudrunを例にして説明しよ
う。

Kudrunは全体の語彙を A.Unversum B. Mensch C. Mensch und Weltとに
分類し、さらにA. Unversum を I. Himmel und Atmosphäre II. Erde III.
Pflanzewelt IV. Tierwelt に、B. Mensch を I. Körperliches Wesen II.
Geist/Seele III. Mensch als soziales Wesen に、C. Mensch und Weltを I.
A Priori II. Wissenschaft und Technik
に分類し、それぞれさらにいくつかの下位分類している。

たとえば A.Universum I. Himmel und Atmosphäreを例に取れば、

a) Himmel und Himmelkörper

sterne

(morgensterne 1/1)

13551

sunne

.

b) Wetter und Winde

erwärmen

7423

.

c) Material

stoup

10192

これは語彙の意味領域を重視したグロサールおよびインデックスといっ
てよく、前の二つのタイプとは基本的にコンセプトが異なっている。騎士
道という大きな概念の語彙を調べたいときは、Ritterwesenを見ればPferd,
Reiten, Waffen, Schlacht, Burg, Turnierwesen等に関する関連語のインデッ
クスを見つけることができる。

また対象となる作品に関して言えば、これらのインデックスやコンコー
ダンツは一つの作品を対象としたもの、ある作家の作品集を対象としたの
ものと、さらにSchweizer Minnesängerや des Minnesangs Frühlingのよう
に様々な作家の作品を対象としたものがある。それは内容や文学ジャン
ルとも関わってくる問題であろう。この中で、注目したいのが、アメリカ
で出版された前述の Clifton Hall の Des Minnesangs Frühling 「ミネザ
ングの春」や Walther von der Vogelweide のコンコーダンツである。た
とえば Des Minnesangs Frühling は次のような構成からなっている。

The KWIC Concordance (見出し語、頻度数、出現箇所の コン
コーダンツ)

The Frequency List (頻度数)

Reverse Index of Form (逆引きインデックス)

Rhyme-Word Index

一冊の本に、1つのコンコーダンツと3つのインデックスが含まれ、用
途によってこれを使い分けできるようになっている。読解の為だけではなく、
様々な中高ドイツ語の言語研究に利用価値が大きいと言えよう。

最後に、場合によっては唯一の道具立てとなり得る一般的な中高ドイツ

語の辞書について若干述べておきたい。中高ドイツ語を読む上で不可欠な辞書として、いわゆる大レクサー(Lexer: *Mittelhochdeutsches Handwörterbuch*)三巻本がある。これは本来、その前に編纂されたBMZ (Benecke, Müller, Zarncke: *Mittelhochdeutsches Wörterbuch*)を補完し、検索しやすくするために作られた辞典であった。しかし、改訂されないまま既にはほぼ100年の月日を経ている。大レクサー編纂以降、多くの作品の校訂本が編纂され、テキストのインデックス、グロサル、コンコーダントの類も数多く作成されていることから、大レクサーの大改訂は必須である。数年前に出版されたFindebuch zum mittelhochdeutschen Wörterbuch¹⁶⁾は大レクサー改訂計画の提唱者であるK. Gärtnerが、大レクサー編纂後に作成されたインデックス類を基にして、大レクサーに加えるべき語彙並びに新しい典拠を示した辞書である。このFindebuchは編者によれば、あくまで大レクサー大改訂の暫定的な仕事であるという。しかし、用例蒐集のために我々にとってより有益な辞書は、やはりBMZであろう。この辞書は中高ドイツ語用例の宝庫であるが、その宝庫は宝の持ち腐れの観が強い。というのもその配列と印刷上に問題があり、せっかくの語義の分類が分かりにくく、検索もしにくい。これを整理し、見やすく再構成すれば、中高ドイツ語の学習がどんなに効率よくなることであろうか。その意味でBMZのNeudruckの際に補遺 (Alphabetischer Index von E. Koller, W. Wegstein und N. R. Wolf Stuttgart 1990) が加わったことはBMZの見出し語の引きにくさ (BMZは語源とその親族関係を基にして見出し語が配列されている) を配慮したもので、それこそ少なくとも見出し語を探すまでの時間は大幅に短縮されたことになる。

16) Gärtner/Kurt u.s.w.: Findebuch zum mittelhochdeutschen Wörterbuch mit einem rückläufigen Index. Stuttgart 1992 このFindebuchの内容については以下の書評参照。

新保雅浩: Findebuch zum mittelhochdeutschem Wörterbuchについて。日本独文学会編 ドイツ文学92号

小レクサー(Lexer:Mittelhochdeutsches Taschenwörterbuch)については改訂がしばしば行われ、また比較的早い時期(1984年)にコンピュータを利用して、この小レクサーの逆引き辞典¹⁷⁾が生まれている。

2. 歴史的ドイツ語研究になぜコンピュータが必要か

これまで、19世紀以来の研究成果とコンピュータを利用した最近の成果を、古いドイツ語のテキストを読むための道具という観点から眺めてきた。コンピュータを利用して作成したインデックスやコンコーダントの類が歴史的ドイツ語の研究分野でも次第に始めていることが分かるであろう。では、なぜコンピュータを利用した作業が必要なのであろうか。

コンピュータの威力は大量のデータ処理とそのスピード、正確な繰り返し機能、さらに検索機能であろう。19世紀には、数十年をあるいは一生をかけて手作業で作成したものが、現代では瞬時に出来上がる。さらに一度データとして入力しておけば、様々な利用の仕方が可能である。それだけではない、手作業より正確である。どんな物を作るにせよ、手作業からコンピュータ利用への移行は当然の成り行きなのであり、これを利用しない手はない。

古いドイツ語研究はコンピュータとは無縁であり、昔ながらの方法で、という考えがあるとすれば、それは妙な先入観に過ぎない。自体はむしろ逆である。限られた文献、数少ない資料の中でともかく読解のための言語的道具を少しでも必要としているのである。テキスト読解の為の道具を合理的に増やすことによって、用例検索や蒐集に費やす時間が大幅に短縮される。特に日本の歴史的ドイツ語の研究分野では、前に述べたとおり、テキスト理解が研究の大半を占めており、このためにあまりに多くの時間が

1994

17) Bachofer/v.Hahn/Mohn: Rückläufiges Wörterbuch der Mittelhochdeutschen Sprache. Stuttgart 1984

費やされてしまい、本来の言語研究以前に、テキスト読解自体が研究そのものとなりかね得ないところがある。そのような状況の中では、読解のために、効率的で、正確で、早い方法を取り入れようとする事は至極もったもなことだと思えるのである。道具立てがたくさんあればあるほど、多くのテキストを早く、正確に読める。

コンピュータ利用の理由はもう一つある。読解した成果を逐一インデックスにその語義や文法形態を書き加えることによって、それ自体がテキストの重要な情報として蓄積され、さらに詳しい道具へと発展する。そして、はじめに入力的设计を立てておけば、入力し終わったデータベースをもとに様々な利用が可能となる。たとえば、Wittenweilerのテキストは語彙の綴りは古典的な中高ドイツ語とはかなり異なる。しかし、この作者の作品の中では当然、綴りの体系がある。テキストがコンピュータに入力されていれば、この綴りの体系を調べることが容易である。体系を知ってしまえば、中高ドイツ語のどの語彙に相当するか迷うことはなく辞書に当たることが出来るのである。まして、ある語句の意味、用法を理解するために、同じ様な文脈の用例が必要である場合は、コンピュータの威力が遺憾なく発揮される。歴史的ドイツ語を読む上で重要な点は、同じ言い回し、用例の頻度が高ければ高いほどその語法が正しい用法であることが確認できるのである。さらに、コンピュータで作成された辞書の補助手段は、歴史的ドイツ語の言語的研究の効率化をも生み出す。ここでは、箇条書きにとどめる。

- ① 言語研究のための用例蒐集
- ② 文法形態に関する活用変化研究のための資料
- ③ 接頭辞、接尾辞研究のための資料
- ④ 出現頻度の統計学的研究の資料
- ⑤ 韻律及び押韻研究のための資料

- ⑥ 語彙の意味論的研究のための資料
- ⑦ 辞書編纂の為の基礎資料
- ⑧ 特定の語彙、特定の品詞 シンタクス上の問題などの用例蒐集
- ⑨ 辞書、文法書にある用例の出典箇所の確認

3. 歴史的ドイツ語研究にどんな道具立てが必要か

では、歴史的ドイツ語の研究を効率化するためにどんな道具が必要かを考えてみよう。これまでの述べてきたことをまとめると、必要な道具については次の二つに大別することができる。

3.1 19世紀以来蓄積された古い文献の改訂作業

前に述べたように、19世紀に手作業で、丹念に時間をかけて作成されたインデックスやグロサル、そのような本は現在絶版であったり、あるいは欠落や誤植があり、さらに何よりも検索に時間がかかる。このような既存のグロサルを新たに新しいデータベースを基にして改訂することが歴史的ドイツ語の読解にとって重要である。ここではその例として、前に述べたP.Piperの古高ドイツ語のOtfrid専用辞書の改訂作業（筆者はすでに具体的に実施しているが）を例にとってみよう。この辞書を次のような形に改訂することが理想である。arbeitという名詞の場合、

見出し語(arabeit) — 強弱変化・性別(stf.) — 語義(Arbeit, Anstrengung, Werk ...)

変化形(arabeiti.) — 数・格形(d. s.) — 出現箇所(5,19,50) — 文脈(sie sint...) を配列し、表示すると、次のようになる。

arbeit stf. Arbeit, Anstrengung, Werk; Trübsal, Bedrängnis; Straf, Qual

arabeiti g.s. 5.,9,34 thiū sin selba guati therā altun arabeiti,

arabeiti d.s	5,19,50	sie sint al ebanreiti	in theru selbun arabeiti.
arabeit a.s.	5,10,5	Ni dua thir ,gaudun,thia arabeit, wanta aband unsih anageit	
	5,23,82	wir mit ginadon thinen	thesa arabeit bimiden !
	5,23,108	wir mit ginadon thinen	thesa arabeit bimiden!
	5,23,118	wir mit ginadon thinen	thesa arabeit bimiden!
	5,23,148	wir mit ginadon thinen	thesa arabeit bimiden !
	5,23,160	wir mit ginadon thinen	thesa arabeit bimiden !
arabeiti n.pl.	1, 18,23	arabeiti manego	sint uns hiar io garawo,
	5,7,52	ni klekent mir zi heiti	thie liebun arabeiti.
	2, 14,4	ni lazent thie arabeiti es frist	tthemo warlicho man ist.
arabeito g.pl.	L48	arabeito ginuag;	mit thulti sama iz ouh
		firdruag;	
	2,14,10	theist dages heizesta	joh arabeito meista.
	5,19,24		
	5,23,11	Biscrimi uns,druhtin guato,	thero selbun arabeito,
	H.113	Chorota er ofto thrato	thero selbun arabeito,
	5,23,79	Biscrimi uns, druhtin guato,	thero selbun arabeito,
	95	Biscrimi uns, druhtin guato,	thero selbun arabeito,
	105	Biscrimi uns, druhtin guatto	thero selbun arabeito,
	115	Biscrimi uns, druhtin guato	thero selbun arabeito,
	145	Biscrimi uns, druhtin guato	thero selbun arabeito,
	157	Biscrimi uns, druhtin guato	thero selbun arabeito,
	161	Therero arabeito	tharbent io ginoto
	5,25,7	...	
	5,25,97	...	

arabeitin d.pl. L.23	Riat got imo ofto in notin,	in suaren arabeitin;
1,18,27	Mit arabeitin werbent	thie heiminges tharbent;
4,7,31	Sie sint thanne in wewen,	in arabeitin seren,
5,6,65	Ist rumo er in then guatin	fon unsen arabeitin,
arabeiti a.pl. L.38	er selbo thulta ouh noti	ju manago arabeiti,
2,16,29	Salig thie in noti	thultent arabeiti,
4,31,8	ja thultist thu zi noti	thio selbun arabeiti.
5,23,9	Ouh zellen thio arabeiti	thie wir hiar thulten noti
2,14,46	theih thuruh thino guati	bimidi thio arabeiti.
4,7,13	Tho zalt in thiu sin guati	thio selbun arabeiti,
4,15,42	arabeiti manago,	thio in tho warun garawo.
2,21,20	thaz thu in thera dati	ni firlasest thie arabeiti.
2,14,110	giangut ir bi noti	in anderero arabeiti.

これに加えて、変化語形から原形を導き出すインデックスがあれば、完璧である。古高ドイツ語のTatian には巻末に語彙と格形の付されたインデックスがあるが、これもコンピュータを用いて整理し、改訂すれば、ずっと検索しやすくなる。Tatianの場合はこのインデックスが唯一のTatian専用辞書となるので、改訂は必須であろう。Tatianの様な散文の場合、問題点としては、グロサールの底本となっているテキストを現存の校訂本とするか、新たにコンコーダントのためのテキストを作るか、という問題が残る。

3.2 新たなインデックス・コンコーダント作成

テキスト専用の辞書、インデックス、コンコーダントがないテキストの為にはまず単純なインデックス作成が急務であろう。その場合、変化してい

る語形から原形を導き出せるようなインデックスが実際の読解の場で有益である。

1. インデックス作成

Wolkensteinの例えばarmを例にとろう。最初の単純なインデックスではarmの変化形がアルファベット順にそのまま配列してある。そのインデックスに品詞と原形を加えた上で、原形見出しの配列にする。

語彙見出し	品詞	原形見出し
-------	----	-------

arm	adj.	arm(adj.)
-----	------	-----------

011/125 volg gruder swester arm und reich

019/033 egen ratt loff arm und reich

.....

arm	sub.	arm(sub.)
-----	------	-----------

001/036 seid mir die bain und arm beslagen sind

001/099 in ainen mit dem tengken arm

.....

arme	adj.	arm(adj.)
------	------	-----------

001/114 und waiss nicht wo mein arme sel hin fert

armen	verb.	armen(verb.)
-------	-------	--------------

043/008 an freuden muss ich armen

armen	sub.	arm(sub.)
-------	------	-----------

016/056 mit armen blanck umbvangen

018/092 das klag ich iren zarten weissen armen

armen	adj.	arm(adj.)
-------	------	-----------

010/036 die got verhengt den armen und den reichen

032//042 und die an hail den armen recht vermeiden

.....

armer	adj.	arm(adj.)
008/001	du armer mensch las dich dein sund hie reuen ser	
035/017	wie ward die nacht mit armer macht so wol bedacht	

2. コンコーダンツ

Wolkensteinのarmの例を再び使えば、次のようになろう。

arme	001/114	und waiss nicht wo mein arme sel hin fert
armen	010/036	die got verhengt den armen und den reichen
	032//042	und die an hail den armen recht vermeiden
.....		
armer	008/001	du armer mensch las dich dein sund hie reuen ser
	035/017	wie ward die nacht mitarmer macht so wol bedacht
.....		
armes	107/009	von trost mich vil armes weib
.....		

armen (verb.) (2)

armen	043/008	an freuden muss ich armen
armt	003/023	armt si an werden oder an gut

最初のアルファベット配列は品詞別原形に分類され、それぞれの変化形が下位分類される。これに語義と文法形態を表示すれば、P.PiperのOtfridグロサールのようになる。この流れを箇条書きにすると、次のようになる。

- 1 総語彙インデックス
- 2 各語彙に原形を加える
- 3 原形を見出し語にし、その下に変化語形の語彙を並べたコンコードナンツ作成
- 4 各語彙に語義を付け、文法形態と頻度数を表示する
- 5 逆引き辞典の作成

すなわちP.PiperのOtfridグロサールとDes Minnesangs Frühlingのコンコードナンツを組み合わせたようなコンコードナンツが歴史的ドイツ語の読解には極めて有効ということになるであろう。

むすび

当面の課題としては、1) すでにある古い文献の改訂作業と、2) ほとんど文献のないテキストに関するインデックス作成作業の着手であろう。別の言い方をすれば、ある特定のテキストのインデックス作成から始まり、ある作家の作品ごとのコンコードナンツ作成、同時代のテキストの辞書へと発展させること、が重要である。古い文献の改訂作業としては、上述の古高ドイツ語Otfridや Tatian以外にはBMZの改訂作業が考慮に入れられる。

我が国（筆者の近辺や筆者自身）でも古いドイツ語テキストに対して個人のレベルあるいはグループでインデックス編纂の試みは始められている。しかし少なくとも古いドイツ語の場合、公開し合い、情報を交換し合って作成した方が本来合理的で能率的であろう。

歴史的ドイツ語の文字状況、韻文か散文か等などの多くの問題点を度外視して、この分野の研究者としてさしあたりどんなものが歴史的ドイツ語テキスト読解の道具として必要で、有効かを述べてみた。コンピュータを利用しての道具立てといっても、あくまでコンピュータが主体ではなく、効率の良い環境の中で人間がゆっくり研究することが主眼であることは言うまでもない。本稿では紙の上の情報を問題にしたが、さらに、多くのデータベースを利用して複数共起する語彙の用例を検索することが手軽にパソコンで可能となれば、これは歴史的ドイツ語の研究にとってかなり有益なことはまちがいないであろう。その辺の可能性をも含めて、どこまで歴史的ドイツ語研究者の要望に応えてくれるかは、コンピュータ技術とその専門家の問題であろう。

第 II 部

高瀬 誠

1. 研究のための資料と道具について

1.1 資料について

コンピュータを利用して、歴史的ドイツ語に関する研究活動を行うために、まず必要となる資料を分類してみると、次の様なものとなる：

1. これまで築き上げられてきた業績資料
2. これからの研究に必要となる新たな資料

このうち1は、これまで有益に利用されてきたものを、コンピュータを利用する事でさらに有益に利用できる可能性を持つ事になる。そして2は、これまで作成される事がなかった様な資料であり、それらはとりわけ、これまであまり研究が進まなかった資料の研究を促進させる事ができよう。またそれによってこれまで培われた業績を補足・修正する事も可能となる。こうした点に関しては既に述べられている事なので、ここでは繰り返さない。

1の資料をデータ化したものには、例えば主として次の様なものが考えられる：

* 写本・印刷本等一次資料のテキスト型データ¹⁾

1) 一次資料には、ここで挙げたテキスト型データにはなりにくい情報も記載されている場合がある。そうしたものは画像の形でデータ化し、分析の対象とする場合も考えられるが、これは本稿の対象からはとりえず外す事にする。しかし、データ化の過程で、データにはならずに残ってしまう情報もあり得る事を忘れてはならない。

* 一次資料を元に作成された校訂本のテキスト型データ²⁾

* 既刊の当該言語の辞書のテキスト型データ

2 に属するデータとしては、次の様なものが挙げられるであろう:

* これまであまり研究の進んでいなかった一次資料のテキスト型データ

* 各作品毎の Index, Konkordanz, 辞書のテキスト型データ

* 各時代・各地域毎の言語資料のテキスト型辞書データ

1.2 道具について

歴史的ドイツ語のテキストを調査・研究するための道具として、これまでも辞書、文法書、校訂本の注釈、Konkordanz, Wordindex, 解説書、翻訳書等が利用されてきた。そうしたもののの中には、辞書や文法書・解説書等の様な、コンピュータを利用する/しないに関らず、調査・研究には必要な道具がある。また辞書の様にコンピュータ上のデータとなっていると、極めて有効に利用できるものもある。

一方コンピュータを利用する事で、従来の用途で使われなくなる可能性のあるものもある。例えば出現している語を単純にアルファベット順に並べた Wordindex の様なものがこれに相当する。コンピュータを利用しなかった、いわば手作業で様々な調査・研究が行われていた頃には有用であったこれらの道具も、コンピュータを利用した調査・研究では、1.1 の 1 に属する 3 種の資料と適切な検索・分析ツールがあれば、それらは不要となる。なぜなら、調査対象になる資料と、検索・分析の条件が一定であ

2) これについても、作品毎にどの校訂本をデータ化するか(あるいはできるか)という問題がある。さらに、その校訂本のどの範囲までをデータ化するのか、といった問題も存在する。

れば、結果は常に一定となり、必要な時に、同じものを、何度でも、短時間に、引き出す事ができるからである³⁾。

1.3 Wordindex, Konkordanz について

歴史的ドイツ語の研究に必要な道具のうち、コンピュータを利用したものの中で、Wordindex や Konkordanz に対しては、とりわけ強い要望があるようである。そこでここでは、この資料について特に述べてみたい。

1.3.1 各種 Index の役割について

まず、各種の Index の役割についてここで確認しておく。Wordindex や Konkordanz が作成される目的と役割には以下の様な点が挙げられよう：

1. 当該語句のいわゆる「原形」や変化系列の判定を容易にする
2. 当該語とその同音異義語の区別を明確にする
3. 当該語句の用例検索を容易にする

これにより、各種のテキスト——とりわけ研究のあまり進んでいないテキスト——の解読が容易になり、当該テキストの Syntax 等の分析研究はもとより、当該言語の体系的研究や同時代の他言語との比較研究が進展する事が期待される。さらに、個別のテキストに対する Index, Konkordanz の記述が、当該の言語を記述する辞書・文法書等の記述へ反映する事で、それらを一層充実・発展させる事も十分に考えらる。つまり各種の Index, Konkordanz は、最終的に辞書や文法書へデータが体系的に集約されるための中間的なデータ集という位置付けもできる。

3) データ量やアルゴリズム、あるいはコンピュータの処理能力等の関係から、検索・分析ツール自体が Index を持つ場合もある。しかしこれは純粋に技術的な要請から出てくるものであり、そこで作成された Index を直接調査・研究のための道具として利用するケースは、特殊なものと考えられる。

1.3.2 コンピュータを利用した簡単な Index の作成例

コンピュータを利用する事で、一般に各種の Index は比較的容易に作成する事ができる。ただし、必要となるデータの形式や作成手順、必要な手間等は、求めている Index の形式によって、変わってくる。

よく求められる理想的な Index の形式をここで想定してみると次の様になろう：

＊見出し語 (テキストで使われている語のいわゆる「原形」)

この下には次の様な要素が記述される：

1. 品詞
2. 変化形と文法的特性
3. 使用されている位置表示記号と文例
4. 使用頻度

このうち、各変化形と位置表示記号・文例については、もとになるテキスト資料のデータがきちんとしていれば、かなり容易に Index への処理が可能である。例えば、3.2 で示す実例で用いている Hartmann von Aue の „Der arme Heinrich“ のデータ⁴⁾と、Perl というテキスト処理ツールとそのスクリプト (ここでは w-index.pl というファイルになっている)、それに sort コマンド⁵⁾を利用してみると、例えば次の様な事が可能である。

```
/home/takase_m>./w-index.pl heinr.txt | sort -dfb
```

4) [16]に基づいて作成されたデータ。3.2以降で実例として扱うものと同じ。

5) sort というコマンドは、計算機が持つ文字コードの序列で並べ替えを行うツールである。本来であればここでは、各言語にふさわしい序列での並べ替えが求められるが、今の所そうしたツールは残念ながら存在していない。しかし、そうしたツールは今後必ず必要となってくる事から、筆者の所でも目下設計中ではある。

ab	89	ab besten werdekeit,
ab	117	ab s^{i}ner besten werdekeit
abe	257	alsus tet er sich abe
abe	417	er wirfet diu ougen abe mir.
abe	540	daz ir muoter dar abe
abe	620	und ouch daz guot nie abe gebrach.
abe	792	und der w^{ac} abe tweht,
abe	1106	daz si sichs abe t{ae}te.
abe	1191	abe ziehen diu kleit.
aber	195	sprach der meister aber d^{o}:
aber	200	nu enist aber nieman s^{o} r^{i}ch
aber	215	sprach der meister aber d^{o}:
aber	487	d^{o} ir vater aber tete
aber	518	si bereite aber ein bat
aber	756	n^{u} gesw^{i}ge wir aber der n^{ot},
aber	1104	vil tiure wart si aber besworn,
Absal^{one}	85	als ouch an Absal^{one},
adamas	62	st{ae}ter triuwe ein adamas,
ahte	170	s^{i}n gedinge und s^{i}n ahte.

以下省略

1.3.3 出力結果に関する問題点

この様に、オリジナルのテキストデータから、各語形、当該語のある行番号、当該箇所テキストを簡単に出力する事はもちろん、その当該箇所を上例の様に KWIC (Key Word In Context) の形式で出力させる事も比較的簡単にできてしまう。また、使用頻度についても、各語の例文の件数を

計測させる事により、これに加える事は比較的容易である。

しかし、先の Index のモデルのうち、見出し語となるいわゆる「原形」と、各語の品詞・ならびに当該変化形の文法的特性は、ここで利用している資料テキストデータからだけでは、簡単に導き出す事はできない。この問題をクリアするには、主として次の様な方策が考えられる：

1. テキスト資料の文例から、変化形のいわゆる原形たり得るものを捜し出し、変化系列も同様に文例から導き出す
2. 各語形の見出し語となるべき形と文法的特性をあらかじめテキスト資料のデータに書き加えておく
3. 当該言語に有効な形態素分析ツールを用いる

1 の方式は、例えば動詞であれば、話法の助動詞などと共に使用されている部分から不定詞を抽出したり、名詞であれば 1 格の形で用いられている部分から抽出したりする、といった作業を行う事になる。つまり最終的な判断は結局の所はかなりの部分を人手に頼る事になる。しかし、その全てを手作業で行うよりは、はるかに効率よく作業が進む事は間違いない。これは、あまり研究が進んでいない、あるいはほとんど未研究の資料に対しては、比較的有効な手段であろう。

2 の方式では、分析対象になる資料テキストのデータに、あらかじめ問題となるデータを書き込んでおく事で、抽出を容易にするものである。これは既に研究が進んでいる資料に対してはそれなりには有効となり得るが、そうでない資料に対しては、あまり有効ではない。そうした資料の Index は、そもそも当該語句の文法的特性を効率よく判定していくための材料となる中間的な資料であり、それを作成するのにあらかじめ文法的データを入力するなどというのは、本末転倒である。さらに、そうした文

法上のデータは、後で検索をかける際の障害ともなる⁶⁾。ゆえにこうしたデータは言語資料データには入れるべきではない。

3の方式は、テキスト資料データから、各形態素を自動的に判定し、使用されている変化形、文法的特性、品詞、いわゆる原形、などの要素を表示するツールを利用するものである。一見すると、これがもっとも理想的な様に見える。しかし残念ながら歴史的ドイツ語に関するそのようなツールは今の所存在しない。また存在していても、その分析結果を過信するのはやはり危険な事である。なお、現代日本語用の形態素分析ツールは既に幾つか存在する。ここではそのうちの一つである「茶筌」の名を挙げるに留めておく⁷⁾。

以上の様に、既に存在するコンピュータ上のツールを幾つか組み合わせで利用する事で、比較的簡単にかつ短時間のうちに、それなりの形式の Index を作成できる事は明らかである。しかし Index の種類によっては、全ての作業をコンピュータで自動的に行なおうとするのはやはり無理があり、作業工程によってはどうしても人間の手を煩わせなければならないものもある。そうした場合、原テキストのデータから目的とする Index を作成するために、どのような作業手順・工程が必要となり、そのうちのどの部分をコンピュータで自動化できるのか、どの部分を人間の手で行う必要があるのか、といった点を明らかにした上で作業する必要がある。

1.3.4 Index をめぐる考察

ここで今一度、各種の Index, Konkordanz に関して考えてみたい。各種の Index, Konkordanz には、1.3.1 で挙げた様な利用目的とその効果が考えられる。しかし、冷静に考えてみるとその利用目的のうち、1.3.1 の 3

6) もちろん検索に使うツールにも、スクリプト等のアルゴリズムにも依存する事であるのは言うまでもない。

7) 「茶筌」に関しては、[11] ならびに <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html> を参照されたし。

で挙げられている用例検索の簡易化は、Index をわざわざ作らなくとも、元となるテキストデータと然るべき検索用ツールを利用する事で、事足りてしまう。しかも、必要な部分を、極めて短時間に、何回でも出力させる事ができるのである。またその用例の出力形式に関してみても、従来の各種 Index は一行単位での出力表示しかなく、本当に必要としている部分が、全てすぐさま引き出せて利用できる訳ではない。もちろん、KWIC (Key Word In Context) 形式の出力も、„In Context“ と称しているものの、所詮は一行単位での出力しかしていない⁸⁾。その意味では、用例収集を目的として各種の Index を作成するのは全く無意味となる。それよりも、オリジナルのテキストから Perl のスクリプト等の検索ツールを利用して、場合によっては複数行単位で必要部分を引き出す方が、はるかに効率的に研究活動は進められるであろう⁹⁾。

また、1.3.1 で挙げた 1, 2 の様な、いわゆる「原形」や変化系列の確定、同音異義語の確定という作業のためには、各種の Index の役割は大きいと考えられる。しかしそれらは本来、辞書や文法書が担うべき役割である。そして、各種の Index, Konkordanz を先に述べた通り、辞書や文法書へデータを体系的に集約するための中間的なデータ集と位置付けるのであれば、それを加工・利用して確定された要素は、早急に辞書や文法書に反映させ、役目を終わらせるべきものである。そうした目標のないまま、各種の Index, Konkordanz を作成するのはやはり無意味と言えよう¹⁰⁾。

いずれにせよまずは、分析対象となる写本・印刷本等の一次資料データ

8) 筆者が最初に KWIC 形式を知ったのは、今からおよそ 16, 7 年程前の事である。とあるプログラミングの入門書で見掛けたこの出力形式を、面白いものだとは思ったものの、一行単位での出力にやはり当時も強い不満を覚えたのを今でも記憶している。

9) こうした検索に関しては、後述の 3.2.2 を参照されたし。

10) もちろんその作業が終了したからと言って、各種の Index, Konkordanz が全く不必要になる事はないと思われる。そうした形で完備された辞書や文法書などと共に、計算機上に残された各種の Index, Konkordanz は、歴史的ドイツ語のテキストの読解や学習のために、さらには研究作業の諸過程を学習するためにも良き副教材ともなり得る。その意味では、後進の研究者を育成するためにもその存在意義は失われないうであろう。

や、それを元にして作成された校訂本の資料データを、コンピュータ環境に依存しない形できちんと作成し、然るべきツールと共に用意する事こそが、何よりも肝心な事である。

2 データとコンピュータ環境に関する前提

実際の処理について述べる前に、データやコンピュータ環境についての前提条件を整えておく。

2.1 データ

データは、最低限次の様な条件を全て満たし、これに反する特質を持つていてはならない：

1. 統一的に記述が行われている事
2. 特定のコンピュータ環境に依存しない形式になっている事
3. 文字コード等に関して国際的に問題のない形式になっている事
4. データの記述方式はできる限り広く認知されているものである事
5. 誤りのない(もしくは極めて少ない)データである事
6. できる限り誰でもが利用可能な free なものである事

2.1.1 データ型

文書を扱うデータの形式にも一般にバイナリ型のものとテキスト型のものがある。前者は、ワードプロセッサ等により作成されるデータで、特定のコンピュータ環境で扱いやすい反面、データの汎用性を欠いてしまい、他のコンピュータ環境では極めて扱いにくいものとなってしまう。後者は、それに対してエディタ等で作成されるデータで、コンピュータ環境に

依存せず、幅広く利用できるが、既存の文字コードでの表現しかできないという性質も持つ。

本稿で目指している歴史的ドイツ語の調査研究には、特定のコンピュータ環境に依存するようなものは全く無意味である。また、対象としているデータはテキストの形式を取るものである。従って、本稿ではテキスト型のデータを扱う事を前提とする。

2.1.2 特殊文字について

テキスト型データとして歴史的ドイツ語を記述する際に問題となるのが、歴史的ドイツ語特有の文字の処理である。これはコンピュータが持つ文字コードの体系に依存する問題であり、すぐに解決できるものではない。現状では複数の文字コードを組み合わせて、特殊な文字を表現するしかない。以前はそうした文字列パターンをデータ作成者が各者各様に決めてデータ化が行われていた様であるが、それでは様々なデータを統一的に扱う事はできなくなる。

そうした問題を解決するには、各データの記述を統一的に行う必要がある。その記述方法は当然の事ながら、個別の作成者が勝手に決めたものではなく、広く認知され、かつ最も適切な方式で行わねばならない¹¹⁾。

ここでは国際的に広く用いられている組版処理システム $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ 並びに $\mathrm{L}_{\mathrm{A}}\mathrm{T}_{\mathrm{E}}\mathrm{X}$ のドイツ語用スタイルファイル `german.sty` での記述方式を基準とし、そこで定義されているもの以外についても $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ 及び $\mathrm{L}_{\mathrm{A}}\mathrm{T}_{\mathrm{E}}\mathrm{X}$ での標準的記述方式に従う事とする¹²⁾。

11) また将来、現行のものよりも優れた記述方式が考案され普及した場合、極めて容易にその方式へ移行できるようにしておく必要もある。そのためにもデータは広く認知され、最適な形式で統一的に記述されていなければならない。もちろんその新しい記述方式も、2.1 で挙げた条件を全て満たしていなければならないのは当然の事である。

12) $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ 並びに $\mathrm{L}_{\mathrm{A}}\mathrm{T}_{\mathrm{E}}\mathrm{X}$ については、[7], [8], [9], [18], [19] 等を、`german.sty` については [20] を参照。また、これによるテキスト型データの記述については、[1], [2], [5] 等を参照されたし。

2.2 コンピュータ環境

コンピュータの世界では、物事の進展のスピードが通常の世界よりも桁違いに速い。少し前には最新鋭だった技術が瞬く間に古いものと化してしまふ事もしばしばである。本稿で述べる情報が古くなるのにも、おそらくさほど時間は掛からない事と思われる（もちろんある意味ではそれが望ましい事でもあるが）。

またここで述べる事項は、コンピュータの世界では既にかなり以前に解決されてしまったテーマがその大半である。そのため既にコンピュータを利用する者達にとっては、目新しいものはほとんどない¹³⁾。

本稿では、筆者が普段利用している UNIX システムのうちの一つである Linux を中心にして¹⁴⁾、様々なデータ処理の例を挙げるが、他のコンピュータ環境でも類似した処理が可能である。また、ここで前提とするコンピュータ環境は、現段階で有用と思われるものに過ぎず、将来さらに有用なものが出現している可能性もある事をあらかじめ記しておく。

3 データ処理の実例

ここでは、辞書をデータ化して利用する事と、後述の検索ツールを用いて、データから必要な語句を検索・抽出する方法とを考えてみる事にする。

13) 反面その処理方法は、コンピュータを利用する際の極めてスタンダードなものと言う事もできる。いずれにせよ、そうした処理を行った資料の分析研究に、少しだけやる気になりさえすれば、極めて短時間でできてしまう Index 作成程度のものしか見られないのは、コンピュータ利用者の一人として率直に言って驚きを隠せない。本来であればもっと先に進んだ研究活動が行われていて然るべきではないだろうか。その点に関しては、我々を含めた研究者一同は、既に単純軽作業の域に入った Index 作成程度の作業で満足などせず、さらに先へ進むために一層の努力が求められる所であろう。

14) 現在は Slackware 3.4 + PJE-0.1 を中心としたシステムとなっている。

3.1 辞書のデータ化 —— BMZ を例として ——

歴史的ドイツ語テキストの調査・研究に対して、もっとも有用な参考資料に辞書と文法書が挙げられよう。ここ数年、特に中高ドイツ語については文法書は既存のものの改訂が重ねられ、より充実した内容となってきた。

しかし辞書に関しては、改訂作業が遅々として進まず、以前より指摘されている欠点をそのまま引きずってきているのが現状である。とりわけ Benecke/Müller/Zarncke ([13], 以下 BMZ) は、その優れた記述にも関わらず、目的の見出し語を探すのにも何かと手間が掛かる上に、まるで人を寄せ付けたくないかの様な組版のために、必要な情報を取り出すのにかなりの時間が掛かってしまっていたのは、筆者だけではないのではないだろうか。

一例として *geben* に関する BMZ の記述を整理し、例文を除外した形のデータを示す。

I, 500, b, 7 GIBE, GAP, G\^ABEN, GEG"EBEN \textit{gebe}

I, 500, b, 19 1. ohne object.

I, 500, b, 29 2. mit blo"sem dative.

I, 500, b, 32 3. mit blo"sem accus.

I, 500, b, 33 a. etwas au"ser sich (Gr. 4, 599)

I, 501, a, 19 b. sich

I, 501, a, 30 4. mit dativ und accus.

I, 501, b, 40 5. mit dativ und partitivem genit.

I, 501, b, 44 6. mit infin.

I, 501, b, 44 a. ohne ze

I, 501, b, 46 b. mit ze

- | | |
|---------------|---|
| I, 502, a, 2 | 7. mit pr"adikativem subst., adj., oder partic. |
| I, 502, a, 3 | a. swer meister oder rihter wirt gegeben |
| I, 502, a, 4 | "uber s\^{i}n gen\^oz \textit{Renner} 530 |
| I, 502, a, 5 | b. der alte wart unschuldic geben |
| I, 502, a, 6 | \textit{f"ur unschuldig erkl"art Bon.} 62, 66 |
| I, 502, a, 6 | (vgl. oben 3,b) |
| I, 502, a, 12 | 8. mit adv. |
| I, 502, a, 28 | 9. mit pr"apos. |
| I, 502, b, 38 | 10. mit adverb. pr"ap. |

一瞥して明らかなように、この BMZ の記述はかなり体系的なものとなっている。また、ここでは省略してしまったが、この辞書は豊富な例文でも知られているものである。

この BMZ の内容が例文を含めてデータ化され整理されており、Perl などを利用して検索ツールが用意されていて、あたかも CD-ROM 上の現代ドイツ語の辞書と同じように利用できれば、中高ドイツ語テキストの調査・研究に関する時間の大幅な節約が期待できる¹⁵⁾。さらに、BMZ の記述の改訂作業を行う事があれば、コンピュータで扱えるこうしたデータは、その作業を大幅に前進させる事ができるであろう。

テキスト資料のデータ化と共に、これまでの碩学達の手によるこうした辞書記述のデータ化は、コンピュータを利用した歴史的ドイツ語研究にとってばかりでなく、後に続く研究者達がその業績を有効に利用し、発展させていくためにも、極めて重要なものとなってくる。

15) 現代ドイツ語の CD-ROM 化された辞書を利用した経験があれば、応用可能性の大きさは明白であろう。

3.2 資料テキストからの検索・比較

ここでは、様々な資料のテキスト型データから必要とするものを検索・比較する事について述べる。その際実例に利用されるテキスト型データは、筆者により試験的にデータ化された Hartman von Aue 作の „Der arme Heinrich“ [16] を用いる事にする。

なお、データの形式は次の通りとなる。

- * オリジナルテキストの一行がテキストの一行に相当
- * 一行のデータは <tab code><text> とする
- * データは 5 行毎にオリジナルテキストの行番号を付加する
- * その際の一行のデータ形式は <行番号><tab code><text> とする
- * # を行頭に付加した場合はその行はコメント行とする

これにより、当該のデータは次の様なものとなる：

```
Ein ritter s\^o gel\^eret was
daz er an den buochen las
swaz er dar an geschriben vant:
der was Hartman genant,
5  dienstman was er zOuwe.
er nam im manige schouwe
an misl\^{\i}chen buochen:
dar an begunde er suochen
ob er iht des vunde
10 d\^a mite er sw{\ae}re stunde
m"ohte senfter machen,
```

und von s\^o gewanten sachen

以下省略

3.2.1 grep による検索

UNIX での文字列検索は通常 grep というツールがよく使われる。このツールは UNIX の各システムの発展と共に改良が加えられ、egrep や fgrep などの幾つかのヴァリエーションが存在する。ここでは、Linux Slackware 3.4 に収められている GNU grep Version 2.0 (以下では単に grep と記述) を利用する事にする。

grep は、指定したファイル内から正規表現を利用して指定した文字列を検索し、それを含む行を出力する。例えば ritter という語を検索するには、

```
/home/takase_m/>grep ritter heinr.txt
Ein ritter s\^o gel\^eret was
die ein ritter in s\^{i}ner jugent
1340 als ein vrumer ritter sol
```

の様にすれば結果を得る事ができる。しかし、これでは当該行がどこなのか判明しない。さらに複数のデータからの検索なども必要となつてこよう。そこで、GNU grep のオプションスイッチや、ワイルドカードによるファイル名指定などのいくつかの機能を利用して検索すれば、次の様な結果になる:

```
/home/takase_m>grep -4 ritter *.txt
heinr.txt-# 15., neu bearbeitete Aufgabe,
```

heintr.txt-#	besorgt von Gesa Bonath,
heintr.txt-#	Max Niemeyer Verlag, T"ubingen, 1984.
heintr.txt-#	
heintr.txt:	Ein ritter s\^o gel\^eret was
heintr.txt-	daz er an den buochen las
heintr.txt-	swaz er dar an geschriben vant:
heintr.txt-	der was Hartman genant,
heintr.txt-5	dienstman was er zOuwe.
--	
heintr.txt-30	wie ein herre w{\ae}re
heintr.txt-	ze Swaben gesezzen:
heintr.txt-	an dem enwas vergezzen
heintr.txt-	nie deheiner der tugent
heintr.txt:	die ein ritter in s\^{\i}ner jugent
heintr.txt-35	ze vollem lobe haben sol.
heintr.txt-	man sprach d\^o nieman als\^o wol
heintr.txt-	in allen den landen.
heintr.txt-	er hete ze s\^{\i}nen handen
--	
heintr.txt-	si muose iedoch genesen.
heintr.txt-	swaz d\^o scheltens ergienc,
heintr.txt-	der arme Heinrich ez emphienc
heintr.txt-	tugentl\^{\i}chen unde wol,
heintr.txt:1340	als ein vrumer ritter sol
heintr.txt-	dem sch{\oe}ner z"uhte niht gebrast.
heintr.txt-	d\^o der gn\^adel\^ose gast


```
heintr.txt-      s\^{i}ne maget wider kleite
heintr.txt-      und den arz\^at bereite
```

ここでは当該文字列のヒット行と共にその前後 4 行をも出力させているが、それにより当該ヒット行 (上の結果ではファイル名の直後の記号が: になっている部分) が何行目なのか、またおおよそどのような文脈で使われているか、がとりあえず分かるようにはなる。

また、検索文字列に正規表現のメタ・キャラクタを利用する事により、ある語の変化形を一括して検索する事もできる。例えば *helfen* の各変化形を検索するには次のようにする事で可能となろう¹⁶⁾ :

```
/home/takase_m>grep -i -2 'h"\?[aeiuo]lf([est])\)*n*' *.txt
heintr.txt-220      ich enlieze iuch niht verderben.
heintr.txt-      nu enmac des leider niht s\^{i}n:
heintr.txt:      d\^a von muoz iu diu helfe m\^{i}n
heintr.txt-      durch alle n\^ot s\^{i}n versaget.
heintr.txt-      ir m"ueset haben eine maget
--
heintr.txt-      zuo ir kintl\^{i}chen spil:
heintr.txt-      des gap ir der herre vil.
heintr.txt:      ouch half in s\^ere daz diu kint
heintr.txt-      s\^o l\^{i}hte ze wenenne sint.
heintr.txt-335      er gewan ir swaz er veile vant,
--
```

16) ただし、正規表現での記述は何通りかが考えられる場合も多くあり、ここで紹介するものはその一例に過ぎない事を付記しておく。

正規表現については、[2], [3], [6], [10], [12], [14], [21], [22]などを参照。

heinr.txt- iuwer wirt vil guot r\^at.
 heinr.txt- s\^{\i}t ez alsus umbe iuch st\^at
 heinr.txt: daz man iu gehelfen mac,
 heinr.txt-920 ichn ges\^ume iuch niemer tac.
 heinr.txt- herre, ir h\^at uns doch gesaget,
 --
 heinr.txt- als d\^u dir l\^{\i}hte h\^ast ged\^aht.
 heinr.txt- d\^u h\^ast mich des wol innen br\^aht,
 heinr.txt:935 m"ohtest\^u, d\^u h"ulfest mir.
 heinr.txt- des gen"ueget mich von dir.
 heinr.txt- ich erkenne d\^{\i}nen s"uezen muot:
 --
 heinr.txt- Diu maget lachende sprach,
 heinr.txt- wan si sich des wol versach,
 heinr.txt: ir h"ulfe des tages der t\^ot
 heinr.txt-1110 \^uz werlt\^{\i}cher n\^ot:
 heinr.txt- `got l\^one iu, lieber herre,
 --
 heinr.txt-1315 ir w{\ae}ret biderbe unde guot
 heinr.txt- und hetet vesten mannes muot:
 heinr.txt: s\^o helfe mir got, si h\^ant gelogen.
 heinr.txt- diu werlt was ie an iu betrogen:
 heinr.txt- ir w\^aret alle iuwer tage
 --
 heinr.txt- ze jungest gevallen!
 heinr.txt- den l\^on den si d\^a n\^amen,

```
heintr.txt:1520    des helfe uns got. \^amen.
```

```
heintr.txt-
```

この他の `grep` の利用法の詳細は `grep` のマニュアルを参照されたい。

3.2.2 Perl による検索

`grep` による文字列検索は比較的有用なものであるが、問題となる点もある。例えば検索文字列として指定できる正規表現があまり強力なものではない、検索は行単位で行われるのが基本であり、複数行にまたがる検索ができない、等が挙げられる¹⁷⁾。

そうした問題点を避けつつ、より柔軟なテキスト処理が可能なツールが Perl であろう。Perl でも `grep` の様に正規表現を用いた文字列検索ができるが、使用できる正規表現は極めて強力なものとなっている。また Perl はスクリプト言語であるため、スクリプトを記述する事で文字列検索以外にも様々なデータ加工が可能となる。

ここでも検索のために GNU `grep` によく似た結果を出力するスクリプトを試作する事にした。その名前を `search.pl` とする¹⁸⁾。`search.pl` の主な仕様は以下の通りである：

＊起動オプション、検索文字列、検索対象ファイルはコマンドラインから指定

＊検索文字列には Perl で利用できる正規表現が全て利用できるように

17) また正規表現にはメタ・キャラクタが使われるが、UNIX などのシステム上では、利用する shell にも同様にメタ・キャラクタがあり、使用する shell によっては `grep` の様なツールでは、それぞれのメタ・キャラクタとそのエスケープ文字が混合する事で指定が複雑になる、という問題もある。

18) 本稿で利用するスクリプトは全て試作段階のもので、厳密な動作テストを経ていない。そこで本稿では、敢えて当該スクリプトのソースコードを挙げる事はしない事にする。これは 1.3 で扱った `w-index.pl` でも同様である。とりあえずこれらは、必要な結果を出力するブラックボックスの様な道具と御理解頂きたい。また、Perl については、[10]、[21]、[22] 等を参照。

する

* 出力形式は GNU grep のそれを手本としながら、それを補う形とし、以下の通りとする:

1. 出力形式は<ファイル名><ヒットマーク><行数番号><当該行テキスト>とする
2. ヒットマークは検索文字列を含む行では: に、そうでない行は - とする
3. 出力する各行に対して行数番号を加える事で Text 内の位置を明確にする
4. オプションにより、検索文字列を含む行の前後の任意の行数分のデータも出力させるようにする (その際のオプション指定方法は GNU grep の場合と同じにする)

正規表現を用いた Perl での文字列検索の大きな特徴の一つに、複数行にわたる文字列の検索がある。ここではその一例として、(al)sô ... daz ... という語句とその前後 3 行を、„Der arme Heinrich“ のテキストデータから、上記の仕様を持つ search.pl を使って検索してみる。

UNIX 系のシステムではこの search.pl をコマンドラインから次のように起動する事により検索が行える:

```
home/takase_m>./search.pl -n -3 '(al)?s\\^o\\b.*\\ndaz' *.txt
heintr.txt: 1   Ein ritter s\\^o gel\\^eret was
heintr.txt: 2   daz er an den buochen las
heintr.txt- 3   swaz er dar an geschriben vant:
heintr.txt- 4   der was Hartman genant,
heintr.txt- 5   dienstman was er zOuwe.
```

--

heintr.txt- 9 ob er iht des vunde
 heintr.txt- 10 d\^a mite er sw{\ae}re stunde
 heintr.txt- 11 m"ohte senfter machen,
 heintr.txt: 12 und von s\^o gewanten sachen
 heintr.txt: 13 daz gotes \^eren t"ohte
 heintr.txt- 14 und d\^a mite er sich m"ohte
 heintr.txt- 15 gelieben den liuten.
 heintr.txt- 16 nu beginnet er iu diuten

--

heintr.txt- 209 und iuwer reht brechen
 heintr.txt- 210 und dar zuo versprechen
 heintr.txt- 211 beidiu m\^{\i}n silber und m\^{\i}n golt,
 heintr.txt: 212 ich mache iuch mir als\^o holt
 heintr.txt: 213 daz ir mich harte gerne ernert.'
 heintr.txt- 214 `mir w{\ae}re der wille unerwert'
 heintr.txt- 215 sprach der meister aber d\^o:
 heintr.txt: 216 `und w{\ae}re der arzen\^{\i}e als\^o
 heintr.txt: 217 daz man si veile vunde
 heintr.txt- 218 ode daz man si kunde
 heintr.txt- 219 mit deheinem liste erwerben,
 heintr.txt- 220 ich enlieze iuch niht verderben.

--

以下省略

この例では、ファイル名の直後の記号が: になっている箇所が当該文字

列を含む行である。この種の検索はこの他にも例えば da(r) ... mite や beidiu ... und ... などの様な表現形式で、複数行にわたっているものの検索にも有効となろう。

正規表現による検索機能を用いる事で、語形変化をしているある語の変化形全てを検索する事ももちろん可能である。例えば helfen の変化形を検索する場合は、このスクリプト search.pl でも以下の様にできる:

```
/home/takase_m>./search.pl -n -3 'h([eiaou])"?u)lf[est]*n*' heinr.txt
heinr.txt- 219   mit deheinem liste erwerben,
heinr.txt- 220   ich enlieze iuch niht verderben.
heinr.txt- 221   nu enmac des leider niht s\^{v}n:
heinr.txt: 222   d\^a von muoz iu diu helfe m\^{v}n
heinr.txt- 223   durch alle n\^ot s\^{v}n versaget.
heinr.txt- 224   ir m"ueset haben eine maget
heinr.txt- 225   diu vollen manb{v}re
--
heinr.txt- 330   und daz der maget tohte
heinr.txt- 331   zuo ir kintl\^{v}chen spil:
heinr.txt- 332   des gap ir der herre vil.
heinr.txt: 333   ouch half in s\^ere daz diu kint
heinr.txt- 334   s\^o l\^{v}hte ze wenenne sint.
heinr.txt- 335   er gewan ir swaz er veile vant,
heinr.txt- 336   spiegel unde h\^arbant
--
heinr.txt- 916   `entriuwen, lieber herre m\^{v}n,
heinr.txt- 917   iuwer wirt vil guot r\^at.
```

heintr.txt- 918 s\^{i}t ez alsus umbe iuch st\^at
 heintr.txt: 919 daz man iu gehelfen mac,
 heintr.txt- 920 ichn ges\^ume iuch niemer tac.
 heintr.txt- 921 herre, ir h\^at uns doch gesaget,
 heintr.txt- 922 ob ir hetet eine maget
 --
 heintr.txt- 932 iedoch niht ein senftiu n\^ot,
 heintr.txt- 933 als d\^u dir l\^{i}hte h\^ast ged\^aht.
 heintr.txt- 934 d\^u h\^ast mich des wol innen br\^aht,
 heintr.txt: 935 m"ohtest\^u, d\^u h"ulfest mir.
 heintr.txt- 936 des gen"ueget mich von dir.
 heintr.txt- 937 ich erkenne d\^{i}nen s"uezen muot:
 heintr.txt- 938 d\^{i}n wille ist reine unde guot,
 --
 heintr.txt- 1106 daz si sichs abe t{\ae}te.
 heintr.txt- 1107 Diu maget lachende sprach,
 heintr.txt- 1108 wan si sich des wol versach,
 heintr.txt- 1109 ir h"ulfe des tages der t\^ot
 heintr.txt- 1110 \^uz werltl\^{i}cher n\^ot:
 heintr.txt- 1111 `got l\^one iu, lieber herre,
 heintr.txt- 1112 daz ir mir als\^o verre
 --
 heintr.txt- 1314 ich h\^orte ie die liute jehen,
 heintr.txt- 1315 ir w{\ae}ret biderbe unde guot
 heintr.txt- 1316 und hetet vesten mannes muot:
 heintr.txt: 1317 s\^o helfe mir got, si h\^ant gelogen.

heinr.txt- 1318 diu werlt was ie an iu betrogen:
 heinr.txt- 1319 ir w\^aret alle iuwer tage
 heinr.txt- 1320 und s\^{i}t noch ein werltzage.
 --
 heinr.txt- 1517 als\^o m"uezez uns allen
 heinr.txt- 1518 ze jungest gevallen!
 heinr.txt- 1519 den l\^on den si d\^a n\^amen,
 heinr.txt: 1520 des helfe uns got. \^amen.
 --

また当該行だけを出力させる事もできる:

```

/home/takase_m>./search.pl 'h([eiaio]l"?u)lf[est]*n*' heinr.txt
heinr.txt: 222 d\^a von muoz iu diu helfe m\^{i}n
--
heinr.txt: 333 ouch half in s\^ere daz diu kint
--
heinr.txt: 919 daz man iu gehelfen mac,
--
heinr.txt: 935 m"ohtest\^u, d\^u h"ulfest mir.
--
heinr.txt: 1109 ir h"ulfe des tages der t\^ot
--
heinr.txt: 1317 s\^o helfe mir got, si h\^ant gelogen.
--
heinr.txt: 1520 des helfe uns got. \^amen.
  
```


--

以上のように、Perl の正規表現を用いた文字列検索は極めて強力なものであるが、同時に正規表現が持つ極めて複雑な記述方式が、理解を妨げている事も否めない。こうした問題点も、スクリプトの記述によって改善の可能性もある¹⁹⁾。

3.2.3 資料テキストの比較

歴史的ドイツ語のテキストは、写本などの一次資料データはもちろん、校訂本などでも同一作品に対して複数種が存在することが多い。そうした複数あるテキストの異同を調査する必要も場合によっては出てくる。とりわけ、校訂本作成・改訂等の作業がそれに相当する事であろう。

こうした複数のテキスト型データの異同を比較するツールに diff というものが UNIX 系のシステムでは以前から用いられている。この diff は本来、プログラムの更新の際にソースコードから差分を取るためのもので、その差分データを利用して patch というツールで別の場所にあるソースコードを更新させるのである。

この diff を利用すると、同一作品の複数の資料の比較・検討が、大まかなものではあるものの可能になる。ここでは試みに、これまで取り上げてきた Der arme Heinrich の先頭部分の 15 行を [16] のもの (heintr-a.txt) と [17] のもの (heintr-b.txt) とを用いて、diff の動作の一端を紹介する事にする。

コマンドライン上での指定に対し、次の様な結果が得られる:

19) 例えば、検索文字列をいわゆる原形により指定し、その変化系列をあらかじめ登録した辞書から拾い出し、それをもとにして検索をかけるなどの処理をするようにさせる、というのも一つの手段として考えられる。ただし、検索用サーバ等にはそれなりに負荷がかかる事になってしまうという問題点も出てくる。

```
/home/takase_m>diff -c -C 1 heinr-a.txt heinr-b.txt
```

```
*** heinr-a.txt Wed Nov 18 03:07:02 1998
```

```
--- heinr-b.txt Wed Nov 18 03:05:58 1998
```

```
*****
```

```
*** 4,6 ***
```

```
        der was Hartman genant,  
! 5      dienstman was er zOuwe.  
        er nam im manige schouwe
```

```
--- 4,6 ----
```

```
        der was Hartman genant,  
! 5      dienstman was er ze Ouwe.  
        er nam im manige schouwe
```

```
*****
```

```
*** 8,10 ***
```

```
        dar an begunde er suochen  
!       ob er iht des vunde  
10      d\^a mite er sw{\ae}re stunde
```

```
--- 8,10 ----
```

```
        dar an begunde er suochen  
!       ob er iht des funde  
10      d\^a mite er sw{\ae}re stunde
```

上の例では、!がついた部分が、比較対象となっているものと差異がある行である。

4 結び

これまでみてきた様に、これからの歴史的ドイツ語研究にとって必要な道具は、コンピュータを利用する事により、その大部分が極めて容易にかつ以前に比べてはるかに短時間で準備でき、活用できる事は明らかである。もし問題となる点があるとすればそれは、それを実現するための意欲と実行力、並びに研究に対する明確な展望の有無であろう。

また、個人で研究を行う場合、コンピュータに対する正しい知識と認識も必要となる。しかし仮にそうしたものがなくとも、正しい知識と認識を持つ共同研究者がいれば、それを十分に補う事もできる。その意味では、コンピュータに関する知識は必須とは限らない。むしろ当該研究に対する明確な見通しこそが、より重要になってくる。さらにその場合、共同研究者との十分かつ綿密なコミュニケーションも、研究をきちんと進める上では極めて重要となる²⁰⁾。

本稿で取扱った事例は、様々な理由からとりあえずスタンド・アロンの環境下でのものだけに限定した。しかし、ここで取扱った事例もネットワーク環境に対応させる事は可能である。また、例えば Oxford Text Library や Gutenberg Project 等の様に、ネットワーク経由でテキストを提供する動きもある。

このようなネットワーク経由でのデータ処理に関しては、法的な問題やセキュリティの問題等、解決しなければならない問題が多々ある。しかしそうした問題を何らかの形でクリアし、広くデータを共有できる時が今後

20) 共同プロジェクトが失敗する場合、多々あるその原因の一つとして、構成メンバー同士のコミュニケーション不足が挙げられる。時にはくどいと思われる位の打合せや確認も必要であろう。また、当該プロジェクトに対する明確な見通しの欠如も、失敗の原因となり得る事は明白である。この種の問題については、[15], [23] 等を参照されたし。

くる事も期待したい。

参考文献

- [1] 轡田 收／高瀬 誠:「OCR によるデータベース作成」, 学習院大学
計算機センター年報 Vol. 16, 1995 所収.
- [2] 轡田 收／高瀬 誠:「テキスト型ドイツ語データベースとデータ検
索法について」, 学習院大学計算機センター年報 Vol. 18, 1997 所収.
- [3] 坂本 文:「たのしい UNIX —— UNIX への招待——」, アスキー出版
局, 1991.
- [4] 砂原 秀樹／石井 秀治／植原 啓介／林 周志:「プロフェッショ
ナル BSD」, アスキー出版局, 1994.
- [5] 高瀬 誠:「計算機による欧文テキスト処理について——機種依存し
ない統一的处理についての一考察——」, 学習院大学大学院ドイツ文
学語学研究 第 18 号, 1994 所収.
- [6] 高瀬 誠:「Text-Korpus からの検索について——現代ドイツ語動詞
の検索用正規表現の試み——」, 学習院大学ドイツ文学会 研究論集
1, 1997 所収.
- [7] 中野 賢:「日本語 \LaTeX ブック」, アスキー出版局, 1996.
- [8] 野寺 隆志:「楽々 \LaTeX ・第 2 版」, 共立出版, 1994.
- [9] 藤田 眞作:「 \LaTeX_{ϵ} 階梯」, アジソン・ウェスレイ・パブリッ
シャーズ・ジャパン, 1996.
- [10] 前田 薫／小山 裕司／斉藤 靖／布施 有人:「新 Perl の国へよ
うこそ」, サイエンス社, 1996.
- [11] 松本 裕治／北内 啓／山下 達雄／平野 善隆／今 一修／今村
友明:「日本語形態素解析システム『茶釜』version 1.5 使用説明
書」, 奈良先端科学技術大学院大学 松本研究室, NAIST Technical Re-

port NAIST-IS-TR97007, 1997.

- [12] Aho, A. V./Kernighan, B. W./Weinberger, P.J. 共著/足立 高德 訳：
「プログラミング言語 AWK」, 株式会社 トッパン, 1989.
- [13] Benecke, G. F./Müller, W./Zarncke, F.: *Mittelhochdeutsches Wörterbuch.*
Mit Benutzung des Nachlasses von Georg Friedrich Benecke,
ausgearbeitet von Wilhelm Müller und Friedrich Zarncke. Bd. I: 1854,
Bd. II.1: 1863, Bd. II.2: 1866, Bd. III: 1861, Leipzig, Nachdruck 1966.
- [14] Friedl, Jeffery E. F.: *Mastering Regular Expressions*, O'Reilly & Associates, Inc., 1997.
- [15] Gause, Donald C./Gerald M. Weinberg 共著/木村 泉 訳: 「ライト、
ついてますか —— 問題発見の人間学 ——」, 共立出版, 1987.
- [16] Hartmann von Aue: *Der arme Heinrich*, Herausgegeben von Hermann
Paul, 15., neubearbeitete Auflage, besorgt von Gesa Bonath, *Altdeutsche
Textbibliothek* Nr. 3, Max Niemeyer Verlag, Tübingen, 1984.
- [17] Hartmann von Aue: *Der arme Heinrich*, *Mittelhochdeutscher Text und
Übertragung*, Herausgegeben und übersetzt von Helmut de Boor, Fischer
Taschenbuch 480, Fischer Taschenbuch Verlag, Frankfurt am Main, 1981.
- [18] Knuth, D. E. 著/鷺谷 好輝 訳/斉藤 信男 監修: 「 \TeX ブック」, ア
スキー出版局, 1989.
- [19] Lamport, Leslie 著/Edgar Cooke・倉沢 良一 監訳/大野 俊治・小暮
博道・藤浦 はる美 訳: 「文書処理システム \LaTeX 」, アスキー出
版局, 1990.
- [20] Raichle, Bernd: „Kurzbeschreibung `german.sty` und `ngerman.sty` (Ver
sion 2.5) 8. Juli 1998 (für die Version 2.5e)“.
- [21] Schwarz, Randal L./Tom Christiansen 著/近藤 嘉雪 訳: 「初めての Perl
第2版」, オライリー・ジャパン, 1998.

- [22] Wall, Larry/Tom Christiansen/Randal L. Schwarz 著/近藤 嘉雪 訳:
「プログラミング Perl 改訂版」, オライリー・ジャパン, 1997.
- [23] Weinberg, Gerald M. 著/木村 泉 訳: 「システムづくりの人間学 ——
計算機システムの分析と設計を再考する ——」, 共立出版, 1986.

Möglichkeiten der Computertechnologie für die Mediävistik

— ihre heutige Situation und Ausblicke —

Masahiro Shimbo
Makoto Takase

Ziel der vorliegenden Abhandlung ist es , unter zwei Gesichtspunkten zu betrachten, welche Hilfsmittel der Computer für die Mediävistik bereitstellen sollte und könnte : einerseits vom textphilologischen Gesichtspunkt aus (Shimbo), anderseits vom computertechnischen aus (Takase).

I. Teil

M. Shimbo

Historische Sprachstufen des Deutschen, wie z.B. das Althochdeutsche oder Mittelhochdeutsche, sind , genau genommen, keine normalisierten Einheitssprachen wie etwa das Neuhochdeutsche, sondern Gesamtheiten von vielfältigen Dialekten mit bestimmten Merkmalen. Deshalb fehlen bei den Korpusssprachen wie dem Althochdeutschen und Mittelhochdeutschen normalisierte Wörterbücher oder Grammatiken , auch dabei die Kompetenz eines „native speaker“, was die Untersuchung oder auch die Lektüre beträchtlich erschwert. Die allgemeinen Wörterbücher reichen nämlich für einzelne Texte in der Tat leider nicht aus. Daher braucht man neben dem allgemeinen Wörterbuch Forschungsinstrumente , um einen älteren Text

hinreichend philologisch zu verstehen.

Gebraucht werden also Hilfsmittel, die es erlauben, so viele Belege wie möglich nachzuschlagen und ein genaues grammatisches Verständnis des Textes zu entwickeln : d.h. ein dem Text eigener Wortindex und eine dem Text eigene Konkordanz. Dieses Hilfsmittel dient nicht nur dem Textverständnis im historischen Sprachbereich des Deutschen , sondern auch der Sprachforschung(Bei der Anwendung der Valenztheorie auf die althochdeutsche Syntaxanalyse zieht man z.B. statistische Methoden heran, d.h. Aussagen über die Häufigkeit bestimmter Syntagmen bei einzelnen Verben).

Es gibt zwar solche Wortindexe oder Glossare, die seit dem 19. Jh. von den Sprachforschern in entsagungsvoller , zeitaufwendiger Arbeit hergestellt wurden, aber die meisten davon sind heute entweder vergriffene, oder, wenn lieferbar, doch mit vielen Fehlern behaftet (z.B.P.Piper: Otfrids Evangelienbuch mit Einleitung, erklärenden Anmerkungen ausführlichem Glossar und einem Abriss der Grammatik. 1887).

Die Anwendung der EDV ermöglicht es dagegen , für diese Zwecke eine Materialgrundlage zu erstellen, die manuell nicht mehr erreichbar ist. Deswegen werden seit einiger Zeit unter Anwendung der EDV einige lemmatisierte Konkordanzen oder Wortindexe zu mittelhochdeutschen Texte in Deutschland und vor allem in Amerika veröffentlicht, aber ganz wenige zu althochdeutschen oder frühneuhochdeutschen Literaturdenkmälern.

Deswegen ist es von großer Dringlichkeit,

1. Wortindexe oder auch Glossare, die seit langem korrekturbedürftig sind, mit Hilfe der EDV zu verbessern,

2. alle Wortschätze älterer Texte , zu denen es noch weder Konkordanzen

noch Wortindexe gibt, als Daten in den Computer einzugeben, um dann auf dieser Basis neue Wortindexe u. Konkordanzen zu erstellen.

Meiner Ansicht nach sollte eine solche Konkordanz nicht nur alle Belege (mit grammatikalischen Informationen und der Häufigkeit ihres Auftretens) vollständig erfassen, sondern auch jeweils die Kontexte der Belegstellen erfassen.

II. Teil

M. Takase

Wie im I. Teil beschrieben ist, mangeln jetzt noch die Hilfsmaterialien für die Forschung historischer deutscher Texte. Man kann sie aber viel schneller und leichter erhalten und benutzen, wenn man die betreffenden Texte mit dem Computer verarbeitet¹⁾.

Aber weil die Hilfsmaterialien wie Konkordanz nur zwischenstufige Datensammlungen für die Grammatik und für das Wörterbuch sind, und weil man sie ganz schneller und leichter als früher erhalten kann, sind heutzutage die elektronische Daten der historischen deutschen Originaltexte noch wichtiger. Man kann von ihr die gewünschten Daten mit dem Suchprogramm wie z. B. Perl-Script ganz leicht und einfach erhalten.

Neben dem Originaltext werden auch die Texte der historischen deutschen Wörterbücher (vor allem BMZ. beim Mhd.) nützlicher sein, wenn sie über den Computer — wie die CD-ROM Wörterbücher — zugreifbar sind. Es wäre auch nützlich sein, wenn man alle Texte und Hilfsmaterialien über Netzwerk zugreifen könnte²⁾.

1) Dabei muß man aber auch beachten, daß man nicht alle Informationen ganz automatisch erhalten kann. Bei der Verarbeitung der Daten sollte man ganz klar entscheiden, was man mit dem Computer automatisch machen kann.

2) Dazu muß man viele technische und gesetzliche Probleme überwinden.

Da man mit dem Computer die gewünschten Hilfsmaterialien ganz leichter und einfacher erhalten und benutzen kann, wird man seine Forschung auch leichter entwickeln können. Wenn man auch keine Kenntnisse in der Computertechnik besitzt, kann man sie mit Forschungspartnern gut ergänzen, die gute Kenntnisse darin haben. Dabei wird die Perspektive der Forschung auch noch wichtiger sein.

新保 雅浩（学習院大学教授）

高瀬 誠（学習院大学非常勤講師）